

Improving the quality of health care

Research methods used in developing and applying quality indicators in primary care

S M Campbell, J Braspenning, A Hutchinson, M N Marshall

Before we can take steps to improve the quality of health care, we need to define what quality care means. This article describes how to make best use of available evidence and reach a consensus on quality indicators

This is the second of three articles on research to improve the quality of health care

National Primary Care Research and Development Centre, University of Manchester, Manchester M13 9PL

S M Campbell
research fellow

M N Marshall
professor of general practice

UMC St Radboud, WOK, Centre for Quality of Care Research, (229), Postbus 9101, 6500 HB Nijmegen, Netherlands

J Braspenning
senior researcher

University of Sheffield, Section of Public Health, Sheffield, S1 4DA

A Hutchinson
professor in public health

Correspondence to: S Campbell
stephen.campbell@man.ac.uk

BMJ 2003;326:816-9

Quality improvement is part of the daily routine for healthcare professionals and a statutory obligation in many countries. Quality can be improved without measuring it—for example, by guiding care prospectively in the consultation using clinical guidelines.¹ It is also possible to assess quality without quantitative measures, by using approaches such as peer review, videoing consultations, and patient interviews. Measurement, however, plays an important part in improvement.² We discuss the methods available for developing and applying quality indicators in primary care.

What are quality indicators?

Indicators are explicitly defined and measurable items referring to the structures, processes, or outcomes of care.³ Indicators are operationalised by using review criteria and standards, but they are not the same thing; indicators are also different from guidelines (box 1). Care rarely meets absolute standards,⁵ and standards have to be set according to local context and patient circumstances.^{6 7}

Activity indicators measure how frequently an event happens, such as the rate of influenza immunisation. In contrast, quality indicators infer a judgment about the quality of care provided,⁶ and performance indicators⁸ are statistical devices for monitoring performance (such as use of resources) without any necessary inference about quality. Indicators do not provide definitive answers but indicate potential problems or good quality of care. Most indicators have been developed for use in hospitals but they are increasingly being developed for use in primary care.

Principles of development

Three preliminary issues require consideration when developing indicators. The first is which aspects of care to assess^{w1 w2}: structures (staff, equipment, appointment systems, etc),^{w3} processes (such as prescribing, investigations, interactions between professionals and patients),⁹ or outcomes (such as mortality, morbidity, or patient satisfaction).^{w4} Our focus is on process indicators, which have been the primary object of quality assessment and improvement.^{2 10} The second issue is that stakeholders have different perspectives about quality of care.^{2 w5} For example, patients often emphasise good communication skills, whereas managers' views are often influenced by data on efficiency. It is important to be clear which stakeholder views are being represented when developing indicators. Finally, development of indicators requires supporting infor-

Summary points

Most quality indicators are used in hospital practice but they are increasingly being developed for primary care

The information required to develop quality indicators can be derived by systematic or non-systematic methods

Non-systematic methods are quick and simple but the resulting indicators may be less credible than those developed by using systematic methods

Systematic methods can be based directly on scientific evidence or clinical guidelines or combine evidence and professional opinion

All measures should be tested for acceptability, feasibility, reliability, sensitivity to change, and validity

mation or evidence. This can be derived by systematic or non-systematic methods.

Non-systematic research methods

Non-systematic approaches are not evidence based, but indicators developed in this way can still be useful, not least because they are quick and easy to create. An example includes a quality improvement project based on one case study such as a termination of pregnancy in a 13 year old girl.^{11 12} Examination of her medical records showed two occasions when contraception could have been discussed, and this led to the development of a quality indicator relating to contraceptive counselling.

Systematic, evidence based methods

Whenever possible, indicators should be based solely on scientific evidence such as rigorously conducted (trial based) empirical studies.^{13 14} The better the evidence, the stronger the benefits of applying the indicators in terms of reduced morbidity and mortality. An example of an evidence based indicator is that patients with confirmed coronary artery disease should receive low dose (75 mg) aspirin unless contra-indicated, as aspirin is associated with health benefits in such patients.



Further references are available on bmj.com. These are denoted in the text by the prefix w

Box 1: Definitions and examples of guidelines, indicators, review criteria, and standards

Guideline	Definition Systematically developed statements to help practitioners and patients make decisions in specific clinical circumstances. They essentially define best practice ¹	Example If a blood pressure reading is raised on one occasion, the patient should be followed up on two further occasions within 6 months
Indicator	Retrospectively measurable element of practice performance for which there is evidence or consensus that it can be used to assess quality of care provided and hence change it ⁶	Patients with a blood pressure > 160/90 mm Hg should have their blood pressure remeasured within 3 months
Review criterion	Systematically developed statement relating to a single act of medical care. ⁶ The statement is so clearly defined that it is possible to determine retrospectively whether the element of care occurred ⁴	If an individual patient's blood pressure was > 160/90 mm Hg, was it remeasured within 3 months?
Standard:	The level of compliance with a criterion or indicator ⁶	
Target standard	Set prospectively and stipulates a level of care that providers must strive to meet	90% of practice's patients with blood pressure > 160/90 mm Hg should have their blood pressure remeasured within 3 months
Achieved standard	Measured retrospectively and details whether a care provider met a predetermined standard	80% of practice's patients with blood pressure > 160/90 mm Hg had their blood pressure remeasured within 3 months

Systematic methods combining evidence and expert opinion

Many areas of health care have a limited or methodologically weak evidence base,^{2 6 15} especially within primary care. Quality indicators therefore have to be developed using other evidence alongside expert opinion. However, because experts often disagree on the interpretation of evidence, rigorous methods are needed to incorporate their opinion.

Consensus methods are structured facilitation techniques that explore consensus among a group of experts by synthesising opinions. Group judgments are preferable to individual judgments, which are prone to personal bias. Several consensus techniques exist,^{16–19} including consensus development conferences,^{17 w6} the Delphi technique,^{w7 w8} the nominal group technique,^{w9} the RAND appropriateness method,^{20 w10} and iterated consensus rating procedures (table).²¹

Consensus development conferences

In this technique, a selected group of about 10 people are presented with evidence by interested individuals or organisations that are not part of the decision making group. The selected group discusses this evidence and produces a consensus statement.^{w11} However, unlike the other techniques, these conferences use implicit methods for aggregating the judgments of individuals (such as majority voting). Explicit techniques use aggregation methods in which panellists' judgments are combined using predetermined mathematical rules, such as the median of individual

judgments.¹⁷ Moreover, although these conferences provide a public forum for debate, they are expensive¹⁶ and there is little evidence of their effect on clinical practice or patient outcomes.^{w12}

Indicators derived from guidelines by iterated consensus rating procedure

Indicators can be based on clinical guidelines.^{w13 w14} Review criteria derived directly from clinical guidelines are now part of NHS policy in England and Wales through the work of the National Institute for Clinical Excellence. One example is the management of type 2 diabetes.^{w15} Iterated consensus rating is the most commonly used method in the Netherlands,^{w13 w16} where indicators are based on the effect of guidelines on outcomes of care rated by expert panels and lay professionals.^{w17}

Delphi technique

The Delphi technique is a postal method involving two or more rounds of questionnaires. Researchers clarify a problem, develop questionnaire statements to rate, select panellists to rate them, conduct anonymous postal questionnaires, and feed back results (statistical, qualitative, or both) between rounds. It has been used to develop prescribing indicators.^{w18} A large group can be consulted from a geographically dispersed population, although different viewpoints cannot be debated face to face. Delphi procedures have also been used to develop quality indicators with users or patients.^{w19}

Characteristics of informal and formal methods for developing consensus*

Method	Mailed questionnaires	Private decisions elicited	Formal feedback of group choices	Face to face contact	Interaction structured	Aggregation method
Consensus development conference	No	No	No	Yes	No	Implicit
Delphi technique	Yes	Yes	Yes	No	Yes	Explicit
Nominal group technique	No	Yes	Yes	Yes	Yes	Explicit
RAND appropriateness method	Yes	Yes	Yes	Yes	Yes	Explicit
Iterated consensus rating procedure	Yes	Yes	No	Yes	Yes	Explicit

*Based on Murphy et al.¹⁷

Nominal group technique

The nominal group technique aims to structure interaction within a group of experts.^{16 17 w9} The group members meet and are asked to suggest, rate, or prioritise a series of questions, discuss the questions, and then re-rate and prioritise them. The technique has been used to assess the appropriateness of clinical interventions^{w20} and to develop clinical guidelines.^{w21} This technique has not been used to develop quality indicators with patients, although it has been used to determine patients' views of, for example, diabetes.^{w22}

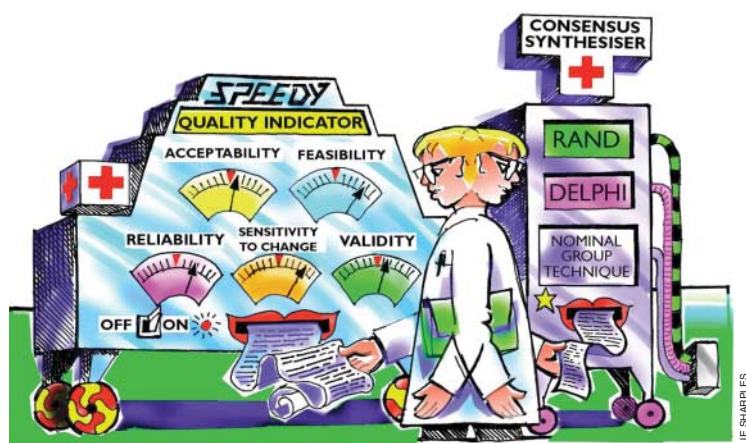
RAND appropriateness method

The RAND method requires a systematic literature review for the condition to be assessed, generation of indicators based on this literature review, and the selection of expert panels. This is followed by a postal survey, in which panellists are asked to read the evidence and rate the preliminary indicators, and a face to face panel meeting, in which panellists discuss and re-rate each indicator.^{w10} The method therefore combines characteristics of both the Delphi and nominal group techniques. It has been described as the only systematic method of combining expert opinion and evidence.^{w23} It also incorporates a rating of the feasibility of collecting data.

The method has been used mostly to develop review criteria for clinical interventions in the United States^{w24} and the United Kingdom.^{7 w25} As with the nominal group technique, panellists meet and discuss the criteria, but because panellists have access to a systematic literature review, they can ground their ratings in the scientific evidence. Agreement between similar panels rating the same indicators has been found to have greater reliability than the reading of mammograms.^{w10} However, users or patients are rarely included, and the cost implications are not considered.

Maximising effectiveness

- Several factors affect the outputs derived using consensus techniques.¹⁹ These include:
- Selection of participants (number, level of homogeneity, etc)
- How the information is presented (for example, level of evidence)
- How the interaction is structured (for example, number of rounds)
- Method of synthesising individual judgments (for example, definition of agreement)



Box 2: Developing and applying review criteria for angina, asthma, and type 2 diabetes

Aim—Quality assessment of angina, asthma, and type 2 diabetes^{9 23}

Sample—60 general practices in England

Patient sample—1000 patients with angina, 1000 with asthma, 1000 with diabetes

Method—Clinical audit; semistructured interviews with nurses and doctors

Acceptability—Used only review criteria that were rated acceptable and valid by the nurses and doctors working in the practices

Reliability—Excluded criteria with an inter-rater reliability κ coefficient < 0.6

Feasibility—Excluded criteria relating to $< 1\%$ of the population sample

- Task set (for example, questions to be rated).

The composition of the group is particularly important. For example, group members who are familiar with a procedure are more likely to rate it higher.^{w26} The feedback provided to panellists is also important.^{w27}

Group meetings rely on skilled moderators and on the willingness of the group to work together in a structured meeting. Unlike postal surveys, group meetings can inhibit some members if they feel uncomfortable sharing their ideas, although panellists' ratings carry equal weight, however much they have contributed to the debate. Panels for group meetings are smaller than Delphi panels for practical reasons.

Research methods for applying indicators

Measures developed by consensus techniques have face validity and those based on rigorous evidence possess content validity. This is a minimum prerequisite for any quality measure. All measures have to be tested for acceptability, feasibility, reliability, sensitivity to change, and validity.^{3 22} This can be done by assessing measures' psychometric properties (including factor analyses), surveys (patient or practitioner, or both), clinical or organisational audits, interviews or focus groups. Box 2 gives an example of the development and testing of review criteria for angina, asthma, and diabetes.^{9 23}

Acceptability—The acceptability of the data collected depends on whether the findings are acceptable to both those being assessed and their assessors. For example, doctors and nurses can be asked about the acceptability of review criteria being used to assess their quality of care.

Feasibility—Information about quality of care is often driven by availability of data.^{w28} Quality is difficult to measure without accurate and consistent information,^{w1} which is often unavailable at both the macro (health organisations) and micro (individual medical records) level.^{w29} Quality indicators must also relate to enough patients to make comparing data feasible—for example, by excluding those aspects of care that occur in less than 1% of clinical audit samples.

Reliability—Reliability refers to the extent to which a measurement with an indicator is reproducible. This

depends on several factors relating to both the indicator itself and how it is used. For example, indicators should be used to compare organisations or practitioners with similar organisations or practitioners. The inter-rater reliability refers to the extent to which two independent raters agree on their measurement of an item of care.²² In one study, five diabetes criteria out of 31 developed using an expert panel⁹ were found to have poor agreement between raters when used in an audit.²³

Sensitivity to change—Quality measures need to detect changes in quality of care in order to discriminate between and within subjects.²² This is an important and often forgotten dimension of a quality indicator.⁶ Little research is available on sensitivity to change of quality indicators using time series or longitudinal analyses.

Validity—Content validity in this context refers to whether any criteria were rated valid by panels contrary to known results from randomised controlled trials.^{w30} The validity of indicators has received more attention recently.^{9 w2 w31} Although little evidence exists of the content validity of the Delphi and nominal group techniques in developing quality indicators,¹⁶ there is some evidence of validity for indicators developed with the RAND method.^{w30} There is also evidence of the predictive validity of indicators developed with the RAND method.^{w32}

Conclusion

Although it may never be possible to produce an error-free measure of quality, measures should be tested during their development and application for acceptability, feasibility, reliability, sensitivity to change, and validity. This will optimise their effectiveness in quality improvement strategies. Indicators are more likely to be effective if they are derived from rigorous scientific evidence. Because evidence in health care is often unavailable, consensus techniques facilitate quality improvement by allowing a broader range of aspects of care to be assessed and improved.⁷ However, simply measuring something will not automatically improve it, and indicators must be used within quality improvement approaches that focus on whole healthcare systems.²⁴

Competing interests: None declared.

- 1 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317-22.
- 2 Donabedian A. *Explorations in quality assessment and monitoring*. Vol 1. *The definition of quality and approaches to its assessment*. Ann Arbor, MI: Health Administration Press, 1980.
- 3 McGlynn EA, Asch SM. Developing a clinical performance measure. *Am J Prev Med* 1998;14:14-21.
- 4 Donabedian A. *Explorations in quality assessment and monitoring*. Vol 2. *The criteria and standards of quality*. Ann Arbor, MI: Health Administration Press, 1982.
- 5 Seddon ME, Marshall MN, Campbell SM, Roland MO. Systematic review of studies of clinical care in general practice in the United Kingdom, Australia and New Zealand. *Quality in Health Care* 2001;10:152-8.
- 6 Lawrence M, Olesen F. Indicators of quality health care. *Eur J Gen Pract* 1997;3:103-8.
- 7 Marshall M, Campbell SM, Hacker J, Roland MO, eds. *Quality indicators for general practice: a practical guide for health professionals and managers*. London: Royal Society of Medicine, 2002.
- 8 Buck D, Godfrey C, Morgan A. *Performance indicators and health promotion targets*. York: Centre for Health Economics, University of York, 1996. (Discussion paper 150.)
- 9 Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK. Development of review criteria for assessing the quality of management of stable angina, adult asthma and non-insulin dependent diabetes in general practice. *Quality in Health Care* 1999;8:6-15.
- 10 Brook RH, McGlynn EA, Shekelle PG. Defining and measuring quality of care: a perspective from US researchers. *Int J Qual Health Care* 2000;12:281-95.
- 11 Pringle M. Preventing ischaemic heart disease in one general practice: from one patient, through clinical audit, needs assessment, and commissioning into quality improvement. *BMJ* 1998;317:1120-4.
- 12 Pringle M. Clinical governance in primary care. Participating in clinical governance. *BMJ* 2000;321:737-40.
- 13 Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Expert consensus on the desirable characteristics of review criteria for improvement of health quality. *Quality in Health Care* 2001;10:173-8.
- 14 McCall A, Roderick P, Gabbay J, Smith H, Moore M. Performance indicators for primary care groups: an evidence-based approach. *BMJ* 1998;317:1354-60.
- 15 Naylor CD. Grey zones in clinical practice: some limits to evidence based medicine. *Lancet* 1995;345:840-2.
- 16 Jones JJ, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376-80.
- 17 Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Ashkam J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;2(3).
- 18 Fink A, Koscoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Pub Health* 1984;74:979-83.
- 19 Black N, Murphy M, Lamping D, McKee M, Sanderson C, Ashkam J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. *Journal of Health Services Research and Policy* 1999;4:236-48.
- 20 Brook RH, Chassin MR, Fink A, Solomon DH, Koscoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *International Journal of Technology Assessment in Health Care* 1986;2:53-63.
- 21 Braspenning J, Drijver R, Schiere AM. *Kwaliteits- en doelmatigheidsindicatoren voor het handelen in de huisartspraktijk*. Nijmegen, Utrecht: Centre for Quality of Care Research, Dutch College of General Practitioners, 2001.
- 22 Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford Medical Publications, 1995.
- 23 Campbell SM, Hann M, Hacker J, Roland MO. Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. *Quality and Safety in Health Care* 2002;11:125-30.
- 24 Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: A framework for change. *Milbank Q* 2001;79:281-315.

The Physician

by Robert Louis Stevenson

There are men and classes of men who stand above the common herd,

the soldier, the sailor, the shepherd not infrequently; the artist rarely; the physician almost as a rule.

He is the flower of our civilisation and, when the stage of man is done and only to be marveled at in history, he will be thought to have shared as little as any in the defects of the period

and most nobly exhibited virtues of the race.

Generosity he has, such as is possible to those who practice an art,

never to those who drive a trade.

Discretion, tested by a hundred secrets, tact tried in a thousand embarrassments, and, what more important, Herculean cheerfulness and courage.

Barbara Jean Graham senior health economist, CSA Information and Statistics Division, Trinity Park House, Edinburgh

Found printed on a scrap of paper dating from around 1925 in the wallet of my grandfather, Dr Eric McKay Reid, with a note scribbled on it from his wife, Barbara Vera Reid, "Can I have this back? It is one of Eric's treasures, B."